DOCUMENT RESUME

ED 265 215 TM 860 082

AUTHOR Avant, Anna H.
TITLE Stability of IQ: A Longitudinal Study of WISC-R Scores of Students Experiencing Learning Difficulties.
PUB DATE Nov 85
NOTE 18p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (14th, Biloxi, MS, November 6-8, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Intelligence Quotient; *Intelligence Tests; *Learning Problems; Longitudinal Studies; Performance Tests; Scores; *Test Reliability; Verbal Tests
IDENTIFIERS *Wechsler Intelligence Scale for Children (Revised)

ABSTRACT

The stability of intelligence test scores over time was examined for the Wechsler Intelligence Scale for Children-Revised (WISC-R). Subjects included 64 children aged 6-16, who had been administered the WISC-R during prior evaluations. These students had been referred because of academic difficulties. One-third of the sample had taken the test twice, one-third three times, and one-third four times. Verbal intelligence quotient (IQ), performance IQ, and full scale IQ scores were available. Correlations and t values values were computed between the first and second, second and third, and third and fourth WISC-R administrations. Results indicated test stability coefficients ranged from .71 to .93 for verbal IQ, from .76 to .88 for performance IQ, and from .69 to .93 for full scale IQ. For the 22 students who had four testing experiences, mean scores were below average each time. Scores generally decreased slightly in subsequent test administrations. An analysis using Pearson correlations showed the following statistically significant correlations between test trials: verbal, .71 to .93; performance, .69 to .88; and full IQ, .66 to .93. It was concluded that the WISC-R is a reliable test, over time, when used with school-age children experiencing learning difficulties. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED265215

STABILITY OF IQ:
A LONGITUDINAL STUDY OF WISC-R SCORES OF STUDENTS EXPERIENCING LEARNING DIFFICULTIES

by
ANNA H. AVANT

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Avant, A. H.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the 1985 Conference of the
Mid-South Educational Research Association
Biloxi, Mississippi

TM 860 082

BEST COPY AVAILABLE

STABILITY OF IQ:

A LONGITUDINAL STUDY OF WISC-R SCORES OF STUDENTS EXPERIENCING LEARNING DIFFICULTIES

INTRODUCTION AND BACKGROUND

Historically, the concept of intelligence has been the focus of on-going debate (Miller & Raynolds, 1983). One feature of intelligence which has received extensive attention is the question of the stability of intelligence over time. As Klonoff (1972) has noted, "The concept of IQ implies constancy and confirmation of constancy can be achieved only by longitudinal studies." In studying the stability of "intelligence," Anastasi (1982) emphasized that, "An important approach to the understanding of the construct, 'intelligence,' is through longitudinal studies of the same individuals over long periods of time." Concern about constancy of IQ arises from the possibility of estimating intelligence from a specific measure obtained at a given point in time. Increasing research on the nature of intelligence reiterates the realization that intelligence is dynamic and variable. Anastasi (1982) noted that, "An extensive body of data has accumulated showing that over the elementary, high school, and college period, intelligence test performance is quite stable" (p. 324).

No longitudinal study has been discovered in the review of the literature in which the Wechsler Intelligence Scale for Children-Revised (WISC-R) has been used exclusively. Studies dealing with the stability of intelligence and the reliability of WISC-R scores have utilized the split-half reliability estimates to determine reliability (Hirshoren, Kavale, Hurley & Hunt, 1977; Dean, 1977; and Mishra & Lord, 1982) or the longitudinal comparisons of the WISC, WPPSI, or WAIS and the WISC-R for selected populations (Thomas, 1980; Solly, 1977; Goodman, 1976; Carvajal, Lane & Gay, 1984; Bishop & Butterworth,

1979; Quattrocchi & Sherrets, 1980). Few studies employing the test-retest procedure were found. Those which were located were two-trial administrations only.

The WISC-R manual provides stability coefficients obtained from a sample of 30 children from six selected age groups in the standardization sample who were retested after an interval of about one month. The average stability coefficients of the IQ scales for this sample were .93, .90, and .95 for the Verbal, Performance, and Full Scale IQs, respectively.

Wechsler (1974) pointed out that a comparison of the mean WISC-R IQs on the first and second testings revealed gains of approximately 3.5 points on the Verbal Scale, 9.5 points on the Performance Scale, and 7 points on the Full Scale.

Tuma & Appelbaum (1980) reported stability coefficients of .95, .89, and .95 for the WISC-R Verbal, Performance, and Full Scale IQs respectively. They obtained increases of one point on the Verbal IQ estimates, and eight and five point increases on the Performance and Full Scale estimates of the IQs of normal children ranging in age from 7.8 to 15.0 years. The mean test-retest interval was 5.84 months. The authors noted that these reliability coefficients were as high or higher than those reported by Wechsler on the standardization sample of ten-year-olds.

Vance, Blixt, Ellis & Debell (1981) compared IQ scores obtained by a sample of 75 learning disabled and retarded children and youth on the WISC-R after a two-year time interval. The stability coefficients revealed for this sample were .80, .91, and .88 for the Verbal, Performance, and Full Scaled IQs, respectively.

Correlations between the first and second administrations of the WISC-R were for Verbal IQ .83, for Performance IQ .84, and for Full Scale IQ .85 for a sample of 30 nine-year-olds with learning difficulties. The mean Verbal IQ

was .40 lower on the second administration while on the Performance and Full Scales, increases in mean IQs of 3.74 and 1.20, respectively, were noted. These correlations and increases in IQ were lower than those reported by Wechsler in his WISC-R manual (Covin, 1977).

Spitz (1983) noted that after an average interval of approximately two years there was a significant average increase of somewhat more than three IQ points in the WISC-R retest, and a substantial correlation of .84. Compared with initial testing, 54% of the 24 students involved in the study deviated by five IQ points or less, 25% by six points, 4% (one student) by nine points, 8% by 11 to 15 points, (both increases), and 8% by 16 points. Most of the changes (75%) were increases. This study was concerned with the Full Scale IQs of mentally retarded students in a relatively short-term residential facility housing primarily mildly and moderately retarded adolescents and young adults. The mean retest intervals ranged from approximately 1.75 to 3.5 years.

The periodic retesting of large numbers of individuals in the public school setting provides the opportunity for the analysis of the stability of the WISC-R for specified populations. The purpose of this study was to determine the degree of consistency and stability between WISC-R scores over time for students referred because of academic difficulties.

METHOD

Subjects

The records of those students tested during the current school year at the Child Guidance Center of a public school system in southern Alabama were searched. This school district consists of rural and urban areas including inner-city schools.

The county served by this school system extends 1,248 square miles. Its population is 364,379, per capital income is $17,011, and there are 91 schools

in the system attended by 66,000 students. Thirty of the schools in the district were represented in this sample. Twelve elementary schools, 11 middle schools, and 7 high schools had students in this sample.

Each of the Verbal, Performance, and Full Scale IQs of each student whose records indicated prior evaluations utilizing the WISC-R was recorded. A table of random numbers was used to select the 64 students for inclusion in the sample. Twenty-two students had been administered the WISC-R four times. Twenty-one students were chosen who had been evaluated three times and 21 who had been evaluated two times with this instrument. This sample of 64 children and youth was comprised of 37 males and 27 females. The racial composition of the sample consisted of 33 blacks and 31 whites. The chronological ages of the subjects ranged from 6 to 16 years.

Materials

The Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) was administered by certified school psychologists. This is an individually administered intelligence test appropriate for students ranging in age from 6 to 16 years. The Verbal Scale consists of the information, similarities, arithmatic, vocabulary, comprehension, and digit span (optional) subtests. The Performance Scale is composed of the picture completion, picture arrangement, block design, object assembly, coding and mazes (optional) subtests. An IQ is computed for each scale as well as a Full Scale IQ. This test is administered in approximately 50 to 75 minutes.

Procedure

Because it is routine to retest students every three years, most scores obtained from the records of previously tested students were those obtained from routine testing by the Child Guidance Department, but a number of scores were available from the WISC-R protocols administered privately by qualified psychologists. Certified school psychologists with extensive experience in

administering and scoring individual tests to subjects of various ages and cultures administered all evaluations effected through the public school system.

Analysis of Data

The Statistical Package for the Social Sciences (Nie, Hull, Jenkins, & Bent, 1975) was used for the data analyses. Correlations and t-values between the first and second, second and third, and third and fourth administrations of the WISC-R comparing the Verbal, Performance, and Full Scale IQs obtained from the test were computed.

RESULTS

An important aspect of a test's reliability is its stability over time. "Indices of consistency or stability might include the following: increments of IQ scores between trials as measured by differences between mean IQ scores, magnitude and direction of change as measured by amount of change in IQ points; and the relationship between repeated test administrations as measured by correlation" (Klonoff, 1972, p. 532). The means, standard deviations, stability coefficients between means of the WISC-R IQ scales, the mean time interval between test administrations, and t-values for the 22 students who received four administrations of this test are summarized in Table 1. This table indicates that the stability coefficients ranged from .71 to .93 for the Verbal IQs, from .76 to .88 for the Performance IQs, and from .69 to .93 for the Full Scale IQs.

Insert Table 1 about here

BEST COPY AVAILABLE

Mean Full Scale IQ scores for the 22 students who were administered the WISC-R four times were below average for each of the four trials; means and standard deviations for the respective trials were 77.18 (17.1), 75.54 (18.58), 76.41 (16.24), and 74.54 (13.96). These results are to be expected in view of the referral basis of the children included in the sample. The means and standard deviations for the respective trials of the Verbal IQs were 78.52 (16.74), 74.81 (18.86), 76.57 (15.34), and 79.57 (15.52). The means and standard deviations for the respective trials for the Performance IQs were 82.19 (18.64), 80.14 (18.37), 79.33 (17.45), and 76.48 (13.20).

The differences between the means for the Verbal IQs were -.71 (a decrease) between Trials 1 and 2; 1.76 between Trials 2 and 3; and 3.00 between Trials 3 and 4. For Performance IQs the differences between the means were -2.05 (a decrease) between Trials 1 and 2; -.81 (a decrease) between Trials 2 and 3; and -2.85 between Trials 3 and 4. The differences between the means for the Full Scale IQs were -1.64 (a decrease) between Trials 1 and 2; .87 between Trials 2 and 3; and -1.87 (a decrease) between Trials 3 and 4. Absolute t-values ranged from .30 for Performance trials 1 and 2 to 1.28 for Performance trials 3 and 4. No t-value was found to be statistically significant at the .05 level of significance.

Table 2 summarizes the data for all students involved in each administration of the WISC-R. The mean Full Scale IQ scores were below average for each of the four trials. Means and standard deviations for each

---

Insert Table 2 about here

---

successive trial were 76.86(16.40), 76.12 (16.96), 75.56 (15.71), and 74.54 (13.96). The means and standard deviations for the respective trials of the

Verbal IQs were 75.52 (16.74), 74.81 (18.86), 76.57 (15.38), and 79.57 (15.52). The means and standard deviations for each successive trials for the Performance IQs were 82.19 (18.64), 80.14 (18.37), 79.33 (17.45), and 76.48 (13.20).

Table 3 presents the Pearson correlations between each trial of the administration of the Verbal Scale. These correlations ranged from .71 between Trials 3 and 4 to .93 between Trials 1 and 2.

Insert Table 3 about here

Table 4 summarizes the correlations between each trial of the administrations of the Performance Scale. These correlations ranged from .69 between Trials 2 and 4 to .88 between Trials 1 and 2.

Insert Table 4 about here

Table 5 indicates the Pearson correlations between each of the Full Scale IQs. These correlations ranged from .66 between Trials 2 and 4 to .93 between Trials 1 and 2.

Insert Table 5 about here

Although previous research indicated that the longer the delay between the original administration and subsequent administrations of the WISC-R the

lower the stability coefficient (Vance et al, 1981), this phenomenon was not substantiated by the present data. The correlation between the first and fourth administrations of the Verbal Scale was .82. A correlation of .79 was noted between the first and fourth administrations of the Performance Scale and the first and fourth Full Scale IQs had a correlation of .76. All correlations were noted to be statistically significant at the .001 level.

DISCUSSION

"Stability of abililty assessment data is essential if placement decisions are not apt to be reconsidered at frequent intervals" (Vance et al, 1981). Because the children referred to school psychologists have most often been identified as having learning difficulties, the stability of IQ estimates for this special population is of the utmost importance. The observed differences in IQ estimates, Verbal, Performance, and Full Scale, when the WISC-R is readministered to a child has implications for making decisions regarding a child's placement and his progress in special education programs. Results from the present study indicate that differences in IQ upon readministration of the WISC-R, thoough present, are not statistically significant. On the average, these changes would rarely lead to erroneous decisions regarding placement.

Frequent assessments of students are desirable and maximize the probabillity of the provision of appropriate services to these students. However, the coefficients obtained on the Verbal, Performance, and Full Scale IQs of the WISC-R and the lack of a significant difference in scores from one administration to the next with the exception of only one set of means suggest that school psychologists can be reasonably certain that a prognosis based on the student's IQ will not change appreciably. The findings of this study provide evidence that the WISC-R is a reliable assessment tool

over time when used with school-aged children and youth who are experiencing learning difficulties.

References

Anastasi, A. (1982). Psychological Testing. New York: Macmillan Publishing Co.

Bishop, D. & Butterworth, G. E. (1979). A longitudinal study using the WPPSI and WISC-R with an English sample. British Journal of Educational Psychology, 49, 156-168.

Carvajal, R.L., Lana, J.M., & Gay, D. (1984). Longitudinal comparisons of Wechsler scales in educable mentally handicapped children and adults. Psychology in the Schools, 21, 137-140.

Covin, T. (1977). Stability of the WISC-R for 9-year-olds with learning difficulties. Psychological Reports, 40, 1297-1298.

Dean, R. S. (1977). Reliability of the WISC-R with Mexican-American children. Journal of School Psychology, 15(3), 267-268.

Goodman, J. (1976). Aging and IQ change in institutionalized mentally retarded. Psychological Reports, 39, 999-1006.

Hirshoren, A., Kavale, K. Hurley, O. & Hunt, J.T. (1977). The reliability of the WISC-R performance scale with deaf children. Psychology in the Schools, 14, 412-415.

Klonoff, H. (1972). IQ constancy and age. Perceptual and Motor Skills, 35, 527-532.

Miller, T. L., & Reynolds, C. R. (Eds.). (1984). Special issue . . . The K-ABC. Journal of Special Education, 18(3).

Mishra, S.P. & Lord, J. (1982). Reliability and predictive validity of the WISC-R with native-American Navajos. Journal of School Psychology, 20(2), 151-153.

Quattrocchi, M. & Sherrets, S. (1980). WISC-R: The first five years. Psychology in the Schools, 17, 297-312.

Solly, D. C. (1977). Comparison of WISC and WISC-R scores of mentally retarded and gifted children. Journal of School Psychology, 15(3), 254-258.

Spitz, H. H. (1983). Intratest and intertest reliability and stability of the WISC, WISC-R, and WAIS full scale IQs in a mentally retarded population. The Journal of Special Education, 17(1), 69-79.

Thomas, P. J. (1980). A longitudinal comparison of the WISC and WISC-R with special education pupils. Psychology in the Schools, 17, 437-441.

Tuma, J. M. & Appelbaum, A.S. (1980). Reliability and practice effects of WISC-R IQ estimates in a normal population. Educational and Psychological Measurement, 40, 671-678.

Vance, H. B. Blixt, S. Ellis, R. & Debell, S. (1981). Stability of WISC-R scores with a sample of exceptional children. Journal of Clinical Psychology, 37, 397-399.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children -Revised. New York: The Psychological Corporation.

Table 1. Summary of Results for each WISC-R Administration (N=22)

Tests	N	Mean IQ	Standard Deviation	Mean Difference	Mean Time Interval	r	t
V1		75.52	16.74				
to	22			-.71	38.23	.93	-.46
V2		74.81	18.86				
to	22			1.76	31.91	.90	.96
V3		76.57	15.34				
to	22			3.00	25.82	.71	1.17
V4		79.57	15.52				
P1		82.19	18.34				
to	22			-2.05	38.23	.88	-1.05
P2		80.14	18.37				
to	22			-.81	31.91	.76	-.30
P3		79.33	17.45				
to	22			-2.85	25.82	.81	-1.28
P4		76.48	13.20				
F1		77.18	17.10				
to	22			-1.64	38.23	.93	-1.10
F2		75.54	18.52				
to	22			.87	31.91	.86	.42
F3		76.41	16.24				
to	22			-1.87	25.82	.69	-.72
F4		74.54	13.96				

Table 2. Summary of Results for each WISC-R Administration (N=64)

Tests	N	Mean IQ	Standard Deviation	Mean Difference	Mean Time Interval	r	t
V1		74.75	15.61				
to	64			.09	38.23	.91	.11
V2		74.84	16.38				
V2		73,38	15,97				
to	43			4.29	31.91	.83	-3.38*
V3		76.57	15.32				
to	22			3.00	25.82	.71	1.17
V4		79.57	15.52				
P1		82.22	17.29				
to	64			-.75	38.23	.90	-.75
P2		81.48	17.35				
P2		79.66	17.17				
to	43			-1.17	31.91	.78	-.69
P3		78.69	16.02				
P3		79.3	17.45				
to	22			-2.86	25.82	.81	-1.28
P4		76.47	13.20				
F1		76.86	16.40				
to	64			-.74	38.23	.93	-.93
F2		76.12	16.97				
F2		74.63	16.45				
to	43			1.93	31.91	.87	1.52
F3		76.56	15.71				
F3		76.41	16.24				
to	22			-1.87	25.82	.72	.48
F4		74.54	13.96				

* $p < .01$.

Table 3. Correlation Matrix of Verbal Scores

Trial	1	2	3	4
1	1.00			
2	.93	1.00		
3	.86	.90	1.00	
4	.82	.82	.71	1.00

Table 4. Correlation Matrix of Performance Scores

Trial	1	2	3	4
1	1.00			
2	.88	1.00		
3	.76	.76	1.00	
4	.79	.69	.81	1.00

Table 5. Correlation Matrix of Full Scale Scores

Trial	1	2	3	4
1	1.00			
2	.88	1.00		
3	.76	.76	1.00	
4	.79	.69	.82	1.00